# Cleanup of Extraneous WorldCat Holdings via SRU

*This document outlines the process used to scan WorldCat holdings for a single holdings symbol and employ an SRU script to check the Alma IZ for holdings corresponding to that symbol. Records in the script output without Alma holdings in a library corresponding to that symbol can be reviewed for batch deletion in WorldCat.*
*The document provides additional instructions on evaluating the deletion candidates and batch-deleting holdings. There is also information on setting holdings on records in batch for those interested in conducting a complete holdings reclamation.*

## Step 1: WorldShare Collection Setup

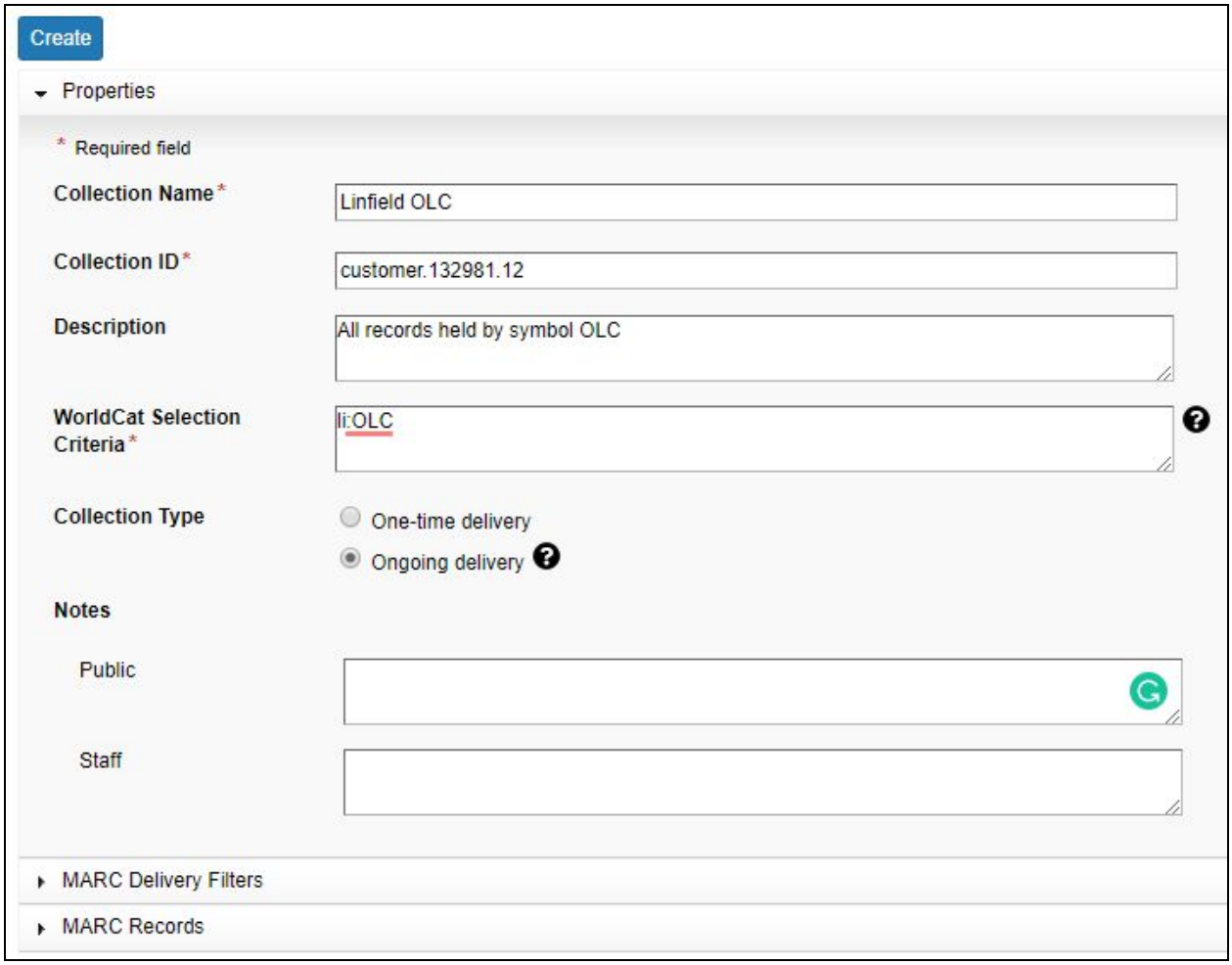
For each holdings symbol you'd like to scan, you need to set up a collection in WorldShare Collection Manager (WSCM). This will enable you to retrieve .mrc files of all holdings associated with that symbol in WorldCat. If your institution has a Cataloging and Metadata subscription with OCLC, then you are able to set up a WSCM instance at no additional cost:

- Call OCLC @ 1-614-793-8682 if you do not have a WorldShare account already set up.
- Ask about setting up a Synchronization Project from ALMA to OCLC using the Product Collection Manager. A Cataloging subscription is required – you can request one if your institution does not have one.
- A username & password will be created for your institution by OCLC, and your WorldShare URL and password will be given verbally and/or by email.
- Once you have this information, log in to WSCM and reset your password.

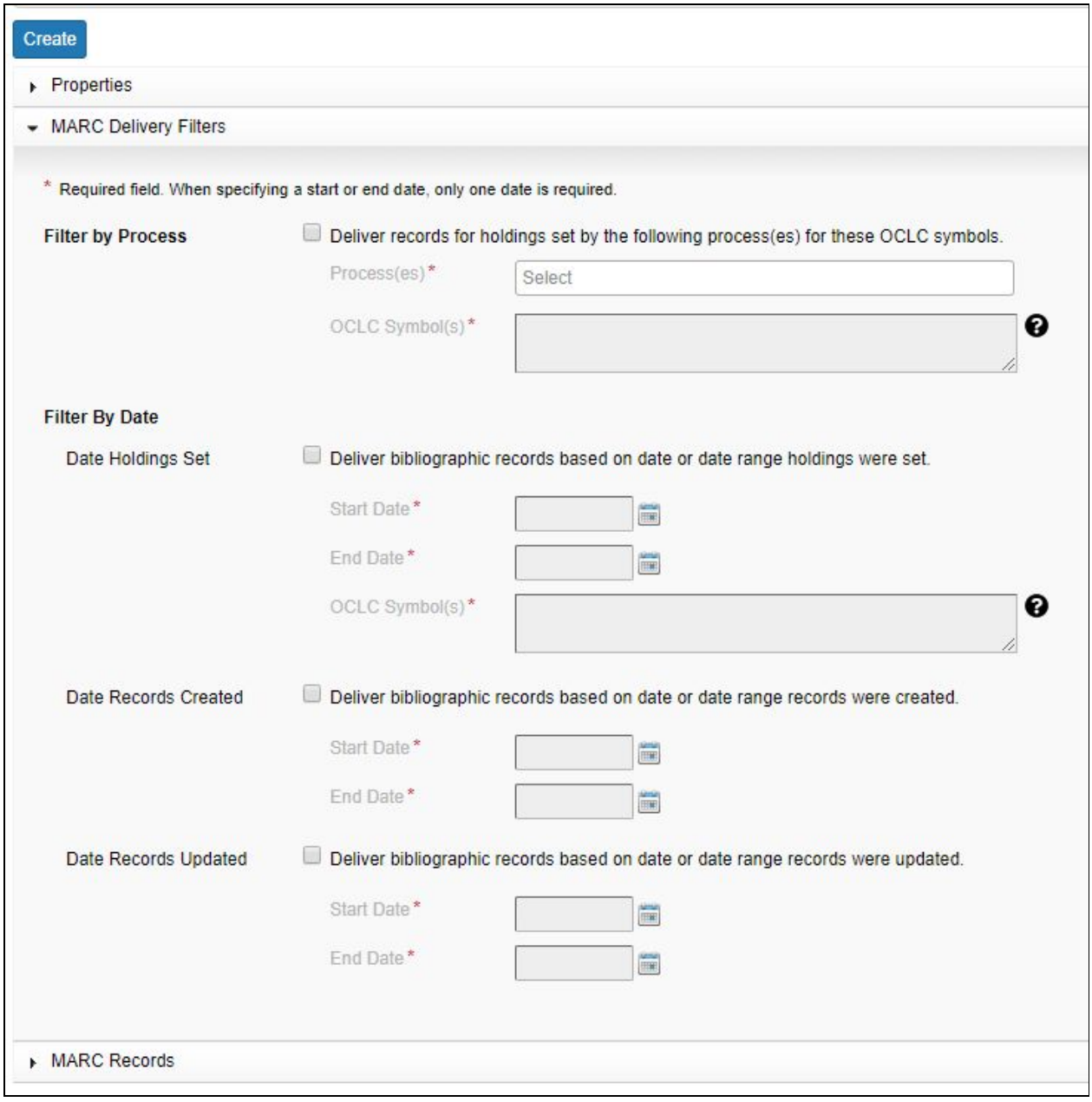
To set up a collection in WSCM for your holdings symbol, log in to WorldShare. Click on Collection Manager to access the search function. Check for existing Collections by leaving the search term box empty and clicking search. If there are no existing collections set up to query your holdings symbol:

- Click Create a Collection
- Collection Type: Query Collection
- Click Create
- Properties Tab:
    - Type in a Collection Name
    - Collection ID will auto-populate
    - Description: optional
    - WorldCat Selection Criteria: li:[*insert your holdings symbol*]

        - Enter only one holdings symbol per collection, or OCLC will deliver a file of deduped records, and you'll be unable to determine whether the correct library has Alma inventory.
    - Collection Type: Ongoing delivery
        - This will allow you to turn record delivery on and off as needed, if you want to re-scan your holdings in the future, or have problems with file delivery.

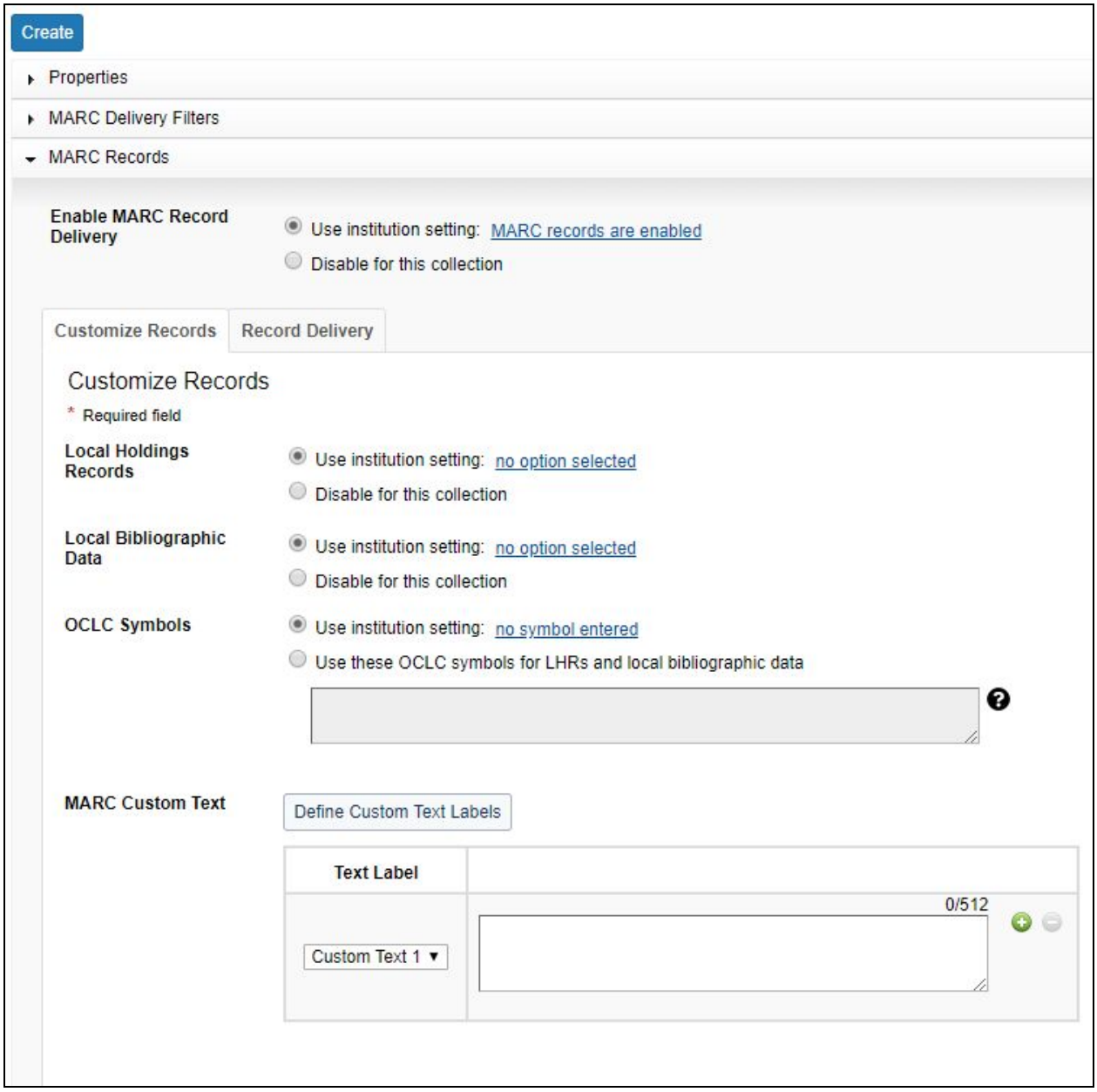



- MARC Delivery Filters Tab:
    - Leave all filters unchecked.



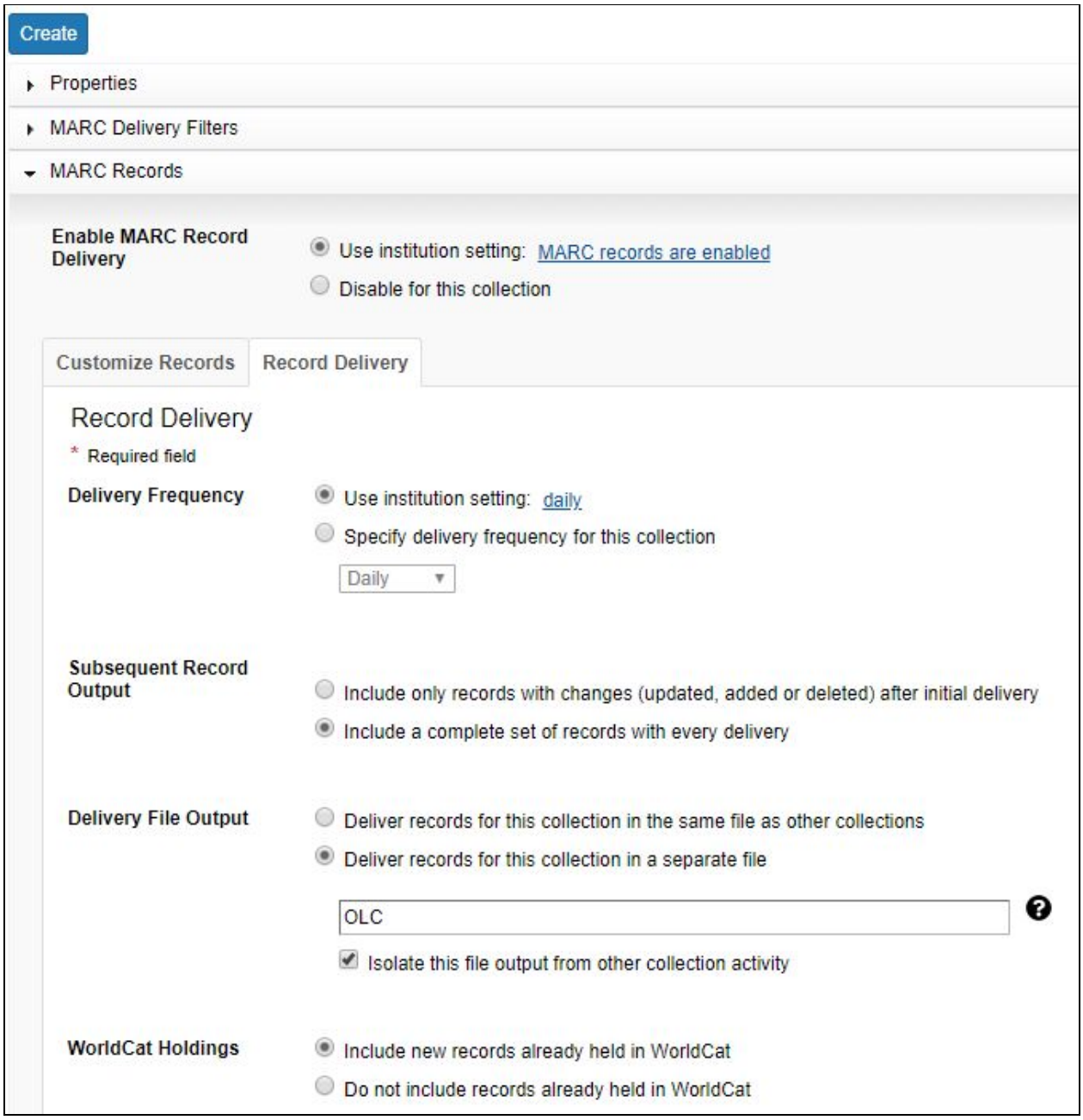

- MARC Records Tab:
    - Enable MARC Record Delivery: Use Institution Setting
        - Institution setting should enable MARC record delivery.
        - Once you have retrieved a set of records for your scan, you can go back into the collection and set this to “Disable for this collection” until you want to run another scan.
    - Customize Records Sub-Tab:
        - Local Holdings Records: No option selected (Institution setting)
        - Local Bibliographic Data: No option selected (Institution setting)
        - OCLC Symbols: No symbol entered (Institution setting)

- MARC Custom Text: None entered




  - Record Delivery Sub-Tab:
    - Delivery Frequency: Choose a frequency that works for you.
      - Daily is good when you're setting up and testing delivery, since you can count on a new file set being generated each day.
    - Subsequent Record Output: Include a complete set of records with every delivery
    - Delivery File Output: Deliver records for this collection in a separate file
      - This prevents your project files from being conflated with other collection deliveries at your institution.

- File Name: [Holdings symbol]
  - This is not the full file name, but a file suffix that will distinguish it from other collection files on the OCLC FTP server.
- Isolate this file output from other collection activity: Yes
- WorldCat Holdings: Include new records already held in WorldCat




- Click Create

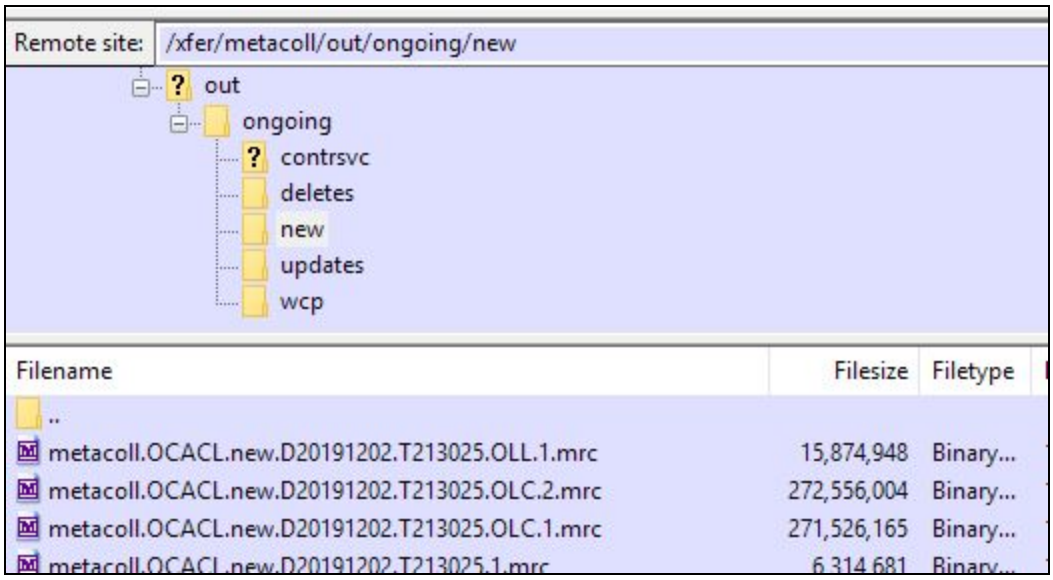
WSCM will give you a record count for your new collection and add it to the list of collections you have activated.

## Step 2: Retrieve and prep OCLC files

Within a day or two of setting up your collection, OCLC will deliver a set of .mrc files containing all of the WorldCat records with holdings set for your symbol to the OCLC FTP server. To retrieve the files, use an FTP client (e.g., FileZilla) to access your OCLC FTP location (OCLC will provide you with this information when you set up your WSCM account.).

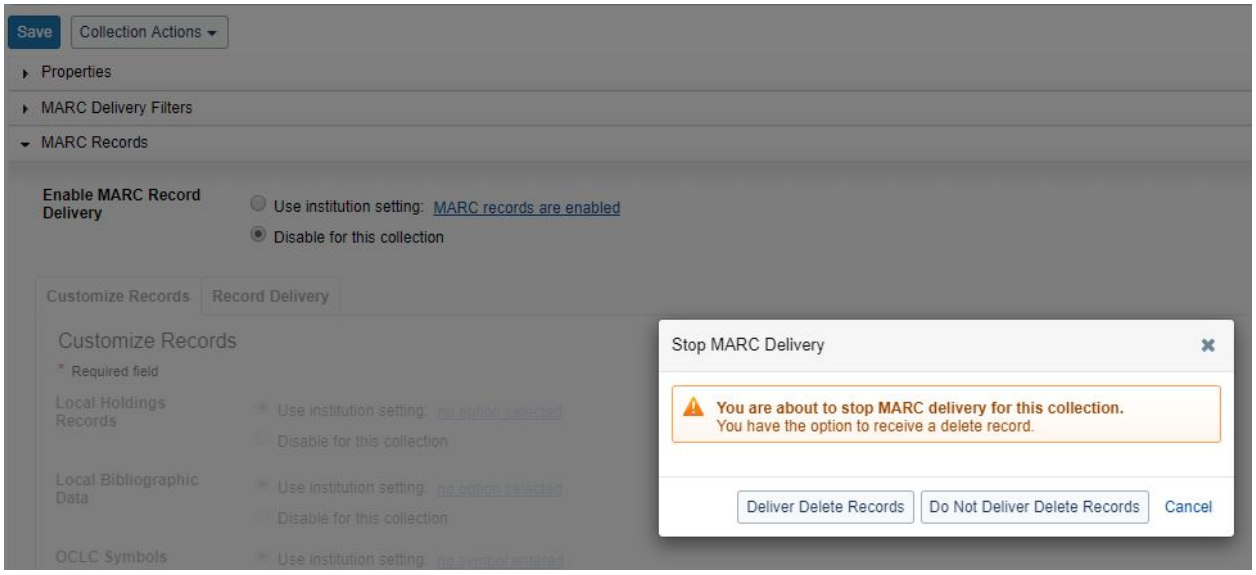
To retrieve your files:

- Connect to the OCLC FTP server via your client of choice.
- Navigate to the /xfer/metacoll/out/ongoing/new directory.
- Look for the file(s) with your holdings symbol in the name.
- Download the file(s).




Once the files are downloaded, you can use MarcEdit's Marc Breaker to get a record count. If the number of records looks right to you, you can go back into WSCM and disable record delivery for the collection. To do this:

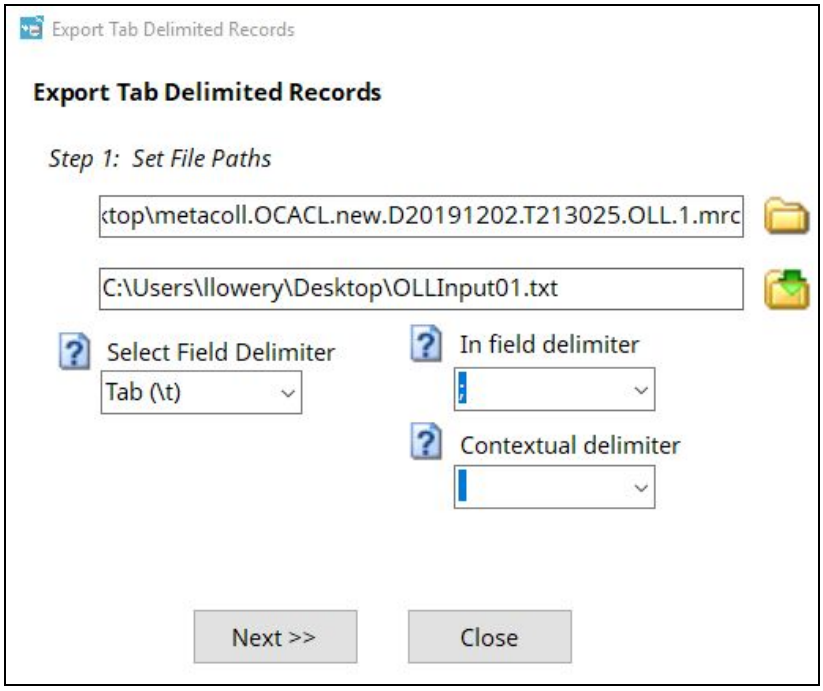
- Log in to WSCM and click search to pull up all of your collections.
- Select the desired collection and navigate to the "MARC Records" tab.
- Change the radio button for "Enable MARC Record Delivery" to "Disable for this collection".
- When prompted, select the "Do Not Deliver Delete Records" option.
- Click Save.




If you didn't get a file, or if you have trouble downloading the file(s), leave record delivery on until you get the problem sorted out with OCLC. Note that if record delivery is left on, existing and merged records will be deposited in the /xfer/metacoll/out/ongoing/updates directory, with new records going to the /new directory and deletes in the /deletes directory. In other words, you will need to check all of the subdirectories under /ongoing/ to get a complete record set after the original file delivery.

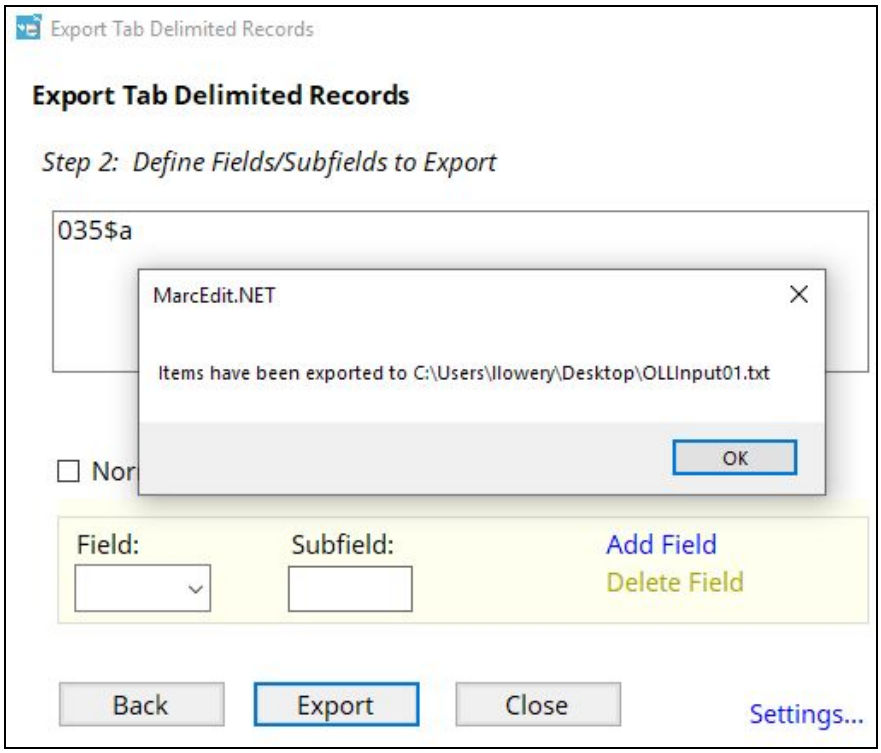
To prep your files:

- Open MarcEdit and use CTRL+T to open the tab-delimited export tool.
- In the first file box, click on the folder icon to browse to your first OCLC file and click Open.
- In the second file box, click on the folder icon to select your destination folder and specify a filename like [*holdings symbol*]input01.txt. Click Save.
- Leave the delimiter settings as-is and click "Next".




- In the Field box, enter 035.
- In the Subfield box, enter “a”.
- Click Add Field.
- Click Export.

When the export completes (this may take a few minutes for large files), you’ll see a note that the export was successful.

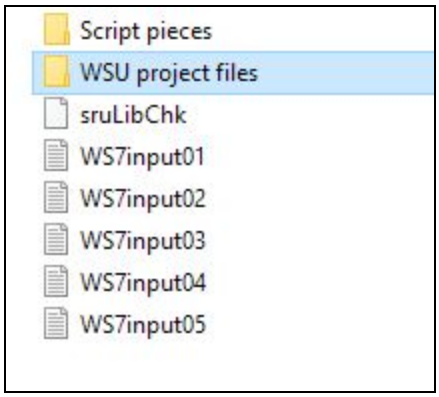


- Repeat these steps for each file of records you downloaded, using the Back button to adjust the input and output filenames so that you have a sequence of files such as “OLLinput01.txt”, “OLLinput02.txt”, etc.
- Once the exports are done, you can archive the .mrc files or delete them.
- Open each text file and remove the “035$a” first line.
- (Optional) Merge the input files into one file if you want to run the script on all of the records in one run. (recommended if you’re running the script on a server)

## Step 3: Prepare and run the SRU script

Before you can run the SRU script on your held record numbers, you need to have the following tools in place:

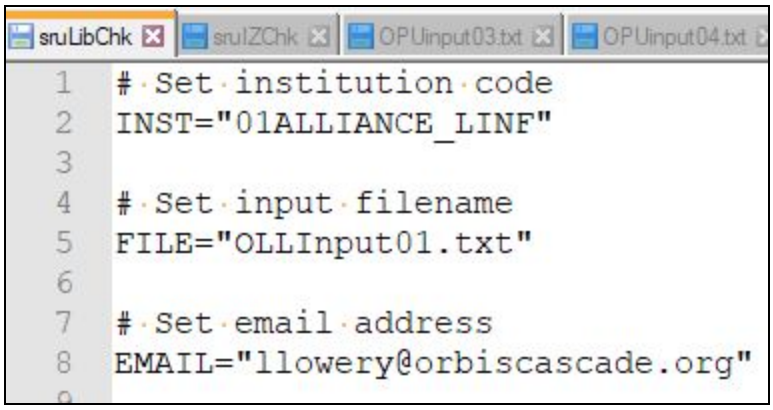
- A workstation or server location with an uninterrupted internet connection and a UNIX distribution that has the xmlstarlet and mailutils packages installed.
    - If you’re using a server and want to start/check the process from different locations, The "screen" utility will allow you to start a process, break out, and disconnect your machine while the program continues running. You can then reconnect later and return to the process to observe its progress. https://linuxize.com/post/how-to-use-linux-screen/ describes how to use the command.
- A copy of the sruLibChk script on that workstation/server - preferably in a working folder set aside specifically for your holding check project.
- Your input .txt file(s), in the same working folder as the sruLibChk script.




To prepare the script:

- Open the sruLibChk script in a text editor (like Notepad++).
- Edit the INST value in line 2 to match the institution code found in your Alma IZ URL.
- Edit the FILE value in line 5 to match your input file name.

- Enter an email address in line 8 where you'd like to receive notifications when the script has finished scanning a file.
- Save the script as type "all files" (don't apply a file extension like .txt!).



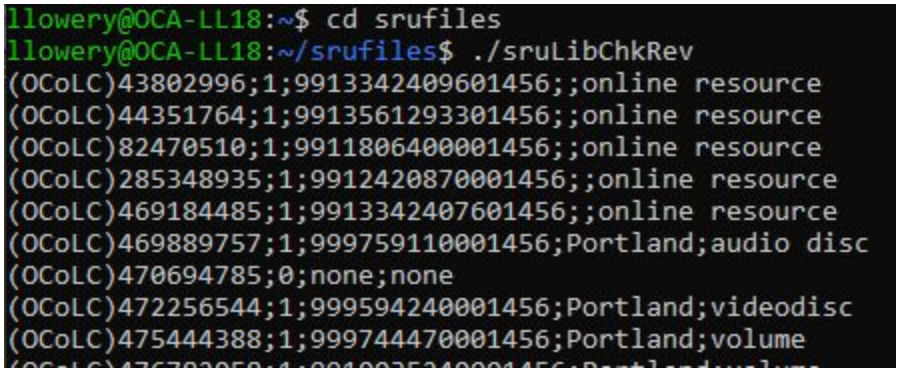

To run the script:

- Open your UNIX command line.
- Navigate to the folder that contains the sruLibChk script and your input file(s).
- Type the command "./sruLibChk" (no quotes) and press Enter.

You should begin seeing output lines in the UNIX terminal window within a few seconds. The same output lines are also added to a new file in your working folder named "liboutput.txt".

## Script output

The sruLibChk script processes a control number roughly every second or so. This means that approximately 67,000 records will be processed per day. This pace is slow, but it ensures that the SRU process does not impact Alma performance negatively.

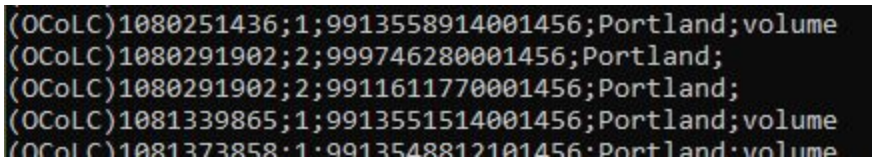



If you are processing multiple files, wait until the script finishes processing the first file, then close your UNIX command prompt, then open the sruLibChk script in your text editor again and update line 5 to designate your next input file. Save the script, then open your UNIX command prompt and to start the next scan as you did before.

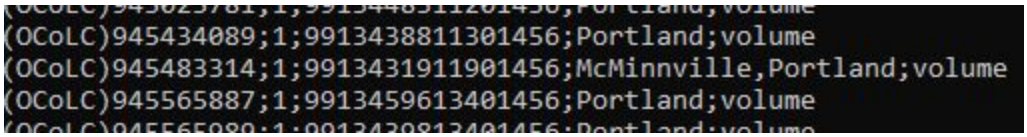
Note that any interruption to internet connectivity or Alma will cause the script to stop running and/or to return “Error” results for control numbers. It is advisable to check the status of the script periodically to ensure that it is running properly and returning valid results.

The script’s output lists the OCLC number, the number of records found in your IZ with that control number, the MMS ID of the Alma record(s), the Holding Library (or Libraries) associated with each MMS ID, and the carrier type for each bib record.

If a control number has multiple records in your IZ, each MMS ID appears on a separate line.



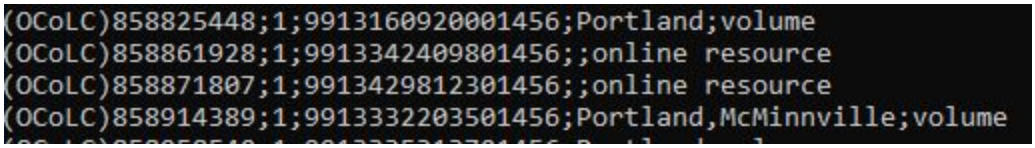

If an Alma record has multiple holdings, the libraries from all holdings records will appear on a single line, separated by commas.



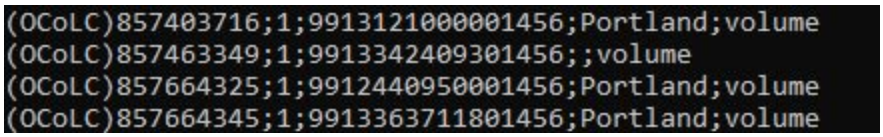

If no Alma record exists for the control number, the script reports 0 records, no MMS ID, and no library:

The carrier is reported to help you disambiguate between electronic titles with no portfolio for a given library and physical resources without an item record in that library.

For example, this capture shows two bib records without library information that have a bib carrier type of “online resource”. Depending on local practice concerning holdings maintenance for electronic resources, these bibs may or may not be candidates for deletion.



In the case below, a physical volume does not have library information in Alma, so it should be reviewed to determine why the bib does not have attached inventory associated with a library, and whether it is a candidate for deletion.




## Step 5: Script output analysis

Once the script has finished processing all of your files, rename the liboutput.txt file (e.g. "[*Holdings symbol*]_Results_01.txt"). Use Excel to open this file, specifying the semicolon as the field delimiter during the import process, and changing the MMS ID column to a format of "Text" (this prevents Excel from putting the MMS ID into scientific notation).

Once the data is in Excel, save the file as an Excel spreadsheet.

- Add Column headings (e.g., "OCN", "# IZ Bibs", "MMS ID", "Libraries", "Carrier")
- For all institutions:
    - Sort on the column with the number of bib records in your IZ.
        - Control numbers that returned errors ("Error" in # IZ Bibs column) can be put back into an input file and re-run. A new output file will be created with the results of these numbers, which can be pasted into your main output spreadsheet.
        - **Control numbers that returned no Alma records (0 IZ bibs) are candidates for batch deletion of holdings in WorldCat.**
        - Control numbers that returned 1 Alma IZ bib can be removed from the list if you're just running an IZ-level scan. If you have multiple holdings symbols, see below for more detailed analysis steps.
        - Control numbers that returned more than one Alma IZ bib should be evaluated as potential duplicate bibs.
            - Note that CZ records in Alma may contribute to the appearance of duplicate records. You can ignore CZ records in your duplicate review.
        - Carrier terms can indicate needed bib record cleanup:
            - Mixed-format carrier types (online resource + physical carrier)
            - Badly-coded carrier terms (e.g., "online", "electronic")
- For institutions with multiple holdings symbols:
    - Using the Data > Filter > Text Filter function on the column of holding libraries:

- Control numbers that return Alma records held by a library corresponding with the holdings symbol processed can be removed from the report as non-problematic.
- Control numbers that returned an Alma MMS ID but no library can be reviewed with the help of the carrier term.
    - These are usually:
        - Electronic resources that can be batch deleted or disregarded, depending on your local practices for electronic holdings, or
        - Bound-with constituent records, which should be reviewed separately for the host bib's presence in the appropriate library.
- **Control numbers that returned holding libraries that do not correspond to the holdings symbol processed are candidates for batch deletion of holdings.**
    - Note that this may also depends on your local preferences for maintaining electronic resource holdings.
    - For example, if you maintain electronic resource holdings for all symbols, but only have a portfolio in your main library, then the online resource entries can be disregarded.

## Step 6: Evaluating the deletion candidates

If you have many candidates for deletion, you may want to review the OCLC records to see if there are any patterns or categories that appear regularly. Such patterns may indicate a local workflow that is unintentionally generating extraneous holdings for your symbol(s). To evaluate the deletion candidates, you can use several tools - Alma, Connexion, MarcEdit, and Excel are the chief tools used.

### Check for false negatives:

Copy the list of candidate OCNs from your output to a new text file and remove the (OCoLC) qualifiers from the numbers. Use this file to create an itemized set in Alma. The set should be empty. If the set has records in it, then remove the OCN for that record from the list of candidates for deletion.

### Check for patterns:

For reviewing the deletion candidates, you have to get the records out of OCLC and (preferably) pull just some of the data out into a spreadsheet. You need Connexion to get the records, MarcEdit to pull selected data out of the records, and Excel to do your comparison. We'll start with Connexion.

**Connexion:**

You can use OCLC Connexion to batch search the OCNs in your list to delete and export a .mrc file of the records' bib data for review. To do that:

- Copy the OCNs from your spreadsheet into a .txt file.
- Log in to Connexion and select **Batch > Enter Bibliographic Search Keys**.
- Select the "OCLC Number" (no:) key from the drop-down list.
- Click **Import** and browse to your text file of OCNs.
- Select the file, and Connexion will import the OCNs. (Don't delete the file of OCNs, since you can re-use it when you batch-delete the holdings.)
- Save the keys and **Close** the window.
- Either empty your local save file, or create a new local save file just for these records.
- Now click **Batch > Process batch**.
- Select the local save file you want the records put into.
- Select the **Online Searches** checkbox.
- Click **OK**. The search will run... and run... and run.
    - Note that there's a limit of 10,000 keys per batch search, and a search that long will take a few hours.
    - You might want to run the batch on a different workstation than the one you're using for cataloging in Connexion, since you can't use your instance for anything else while the search is running.
- Once the search completes, go to **Tools > Options > Export (tab)**.
- Select the File: (Prompt for filename) option and click **Apply**.
    - If no such option exists, click **Create**, and select the **File** radio button.
    - Click **OK**, and set the file type to “All Files”. Select a location and filename on your computer where you'd like the .mrc file of bibs saved. Review the file name and click **OK** when it’s right.
    - If you prefer to be prompted for a file name with each export, click **Cancel** instead of selecting a location/filename.
- **Close** the window.
- Open your local save file, which should now contain all of the records from your list of OCNs.
- Click the first entry to highlight the record.
- Scroll down to the last record in the file, and hold **CTRL+SHIFT** while you click on the last record. This selects ALL of the records in the file.
- Press **F5** to export the records.
- You should be prompted to give the file a name - if you aren't, a new .mrc file with a generic file name should appear in the location you specified under the **Tools > Options > Export (tab)**.

    - If the file is large, it may take a while for the export to complete.
    - My general rule is that when I can drag the Connexion window to a new location on my screen, it's done with the export. (Odd but true!)
- Once the file is exported (check the file with MarcEdit's MarcBreaker tool to make sure the record count is right!), you can use the **Delete** button (or **CTRL+ALT+D**) in Connexion to get rid of all of the records in your save file.
- Don't forget to set your Export back to the usual "Network Zone" setting for day-to-day cataloging when you're done exporting files!

**MarcEdit:**
From there, you use MarcEdit's Tab-delimited export function to pick out which MARC fields/subfields you'd like to examine from each record. That tool creates a .txt file of output that you can input to Excel.

- Open MarcEdit and press **CTRL+T** to open the export function.
- Specify your source file of .mrc records (the file you just exported from Connexion)
- Specify a location and filename where you'd like the results to be saved.
- Click **Next**. (The default delimiter settings are usually sufficient for Excel imports.)
- On the second window, you build a list of the MARC fields and subfields you'd like in your report.
- For each field, enter the MARC tag number and Subfield code, then click **Add Field**.
- Once your list of fields/subfields is complete, click **Export**.
- You'll get a popup note when the export has finished.

From here, you can import that .txt file into Excel and sort/filter/compare as needed.

## Step 6: Deleting the extraneous holdings

Once you have your list of OCNs for holdings deletion together and in a .txt file, you'll use OCLC Connexion to batch delete holdings for the OCNs you identified:

- Connexion batch processes are limited to 10,000 records at a time. If your list of deletions is longer than this, you'll need to break up your list into files of 10,000 OCNs or fewer.
- Log in to Connexion using an account *associated with the symbol you want to remove* from the records.
- Select **Batch > Holdings by OCLC Number**.
- Click **Browse** to select the file of OCNs you want to delete holdings for.
- Select the **Delete Holdings** radio button.
- Click **OK**.

That's it!  The batch will run, removing your symbol's holdings from the designated records. (again, it's good to do this on a separate workstation if you have cataloging to do while it's running).

## Additional reclamation work

It is also possible to use OCLC Connexion's batch holdings feature to conduct the *other* side of a holdings reclamation project - adding holdings to all of the records in Alma that you want holdings set on.  To do this, you'll first need to use Alma search and set features to compile a set (or multiple sets) containing all of the Alma records which should have title-level holdings for your holdings symbol.

- Use Alma's **Export Bibliographic Records** job to push the record metadata out to your local workstation in .mrc binary format.
- Now use MarcEdit's **tab-delimited export** tool to extract the 035$a fields from those records into .txt files.
- It's a good idea at this point to put all the numbers together into one file using Excel and sort them to catch any numbers without the (OCoLC) qualifier.
- Once the numbers are cleaned up, remove the qualifiers and save the file back into .txt format.

Once you have your list together and in a .txt file, you'll use OCLC Connexion to batch set holdings for the OCNs you identified:

- Log in to Connexion using an account *associated with the symbol you want to add to* the records.
- Select **Batch > Holdings by OCLC Number**.
- Click **Browse** to select the file of OCNs you want to set holdings on.
- Select the **Update Holdings** radio button.
- Uncheck the **Set holdings if already held button** - this keeps your batch process from setting holdings on records already held by your symbol.
- Click **OK**.

That's it!  The batch will run (again, it's good to do this on a separate workstation if you have cataloging to do while it's running).

One thing to note is that for a 100% accurate reconciliation of holdings, you'd need to have a cataloging "freeze" for a couple of weeks.  That keeps your "held" records static while you scan for delete candidates via SRU, and keeps the list of Alma records you want to set holdings on static so that you can set holdings on them once and be at "clean slate".